Pertubation de la décomposition spectrale d'une matrice hermitienne

Bernard Philippe

INRIA

CENTRE DE RENNES

IRISA

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél.:(3) 954 90 20

Rapports de Recherche

N° 269

PERTURBATION DE LA DÉCOMPOSITION SPECTRALE D'UNE MATRICE HERMITIENNE

Bernard PHILIPPE

Février 1984

IRISA

INSTITUT DE RECHERCHE EN INFORMATIQUE
ET SYSTÈMES ALÉATOIRES

Campus Universitaire de Beaulieu
Avenue du Général Leclerc
35042 - RENNES CÉDEX
FRANCE
Tél. : (99) 36.20.00
Télex : UNIRISA 95 0473 F

PERTURBATION DE LA DECOMPOSITION SPECTRALE D'UNE MATRICE HERMITIENNE

Bernard PHILIPPE

Publication Interne n° 216
21 pages
Novembre 1983

RESUME :

On présente un algorithme qui permet de déterminer la décomposition spectrale d'une matrice hermitienne à partir d'une décomposition approchée. La méthode est une méthode de correction d'ordre 2 employée itérativement. On montre aussi comment calculer, sans diagonalisation, la matrice $(I + R)^{-1/2}$ lorsque R est une matrice hermitienne de norme inférieure à l'unité.

ABSTRACT :

We present an algorithm which calculates the spectral decomposition of an hermitian matrix from an approximate decomposition. The method is a second order correction method which is used iteratively. We show also, how to calculate the matrix $(I + R)^{-1/2}$, where R is an hermitian matrix whose norm is smaller than one, without diagonalization.

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
(L.A. 227)
UNIVERSITÉ DE RENNES 1 I.N.S.A. DE RENNES

INSTITUT NATIONAL DE RECHERCHE
EN INFORMATIQUE ET EN AUTOMATIQUE
(LABORATOIRE DE RENNES)

IRISA

PUBLICATIONS INTERNES

PERTURBATION DE LA DECOMPOSITION SPECTRALE D'UNE MATRICE HERMITIENNE

Bernard PHILIPPE

PUBLICATION INTERNE No 216

SOMMAIRE

Soit A une matrice hermitienne d'ordre N, dont on connaît la décomposition spectrale : soit X une matrice unitaire telle que la matrice $\Lambda = X^* A X$ soit diagonale.

On se propose alors de déterminer la décomposition spectrale de la matrice $A + \Delta$, où Δ est une perturbation hermitienne "assez petite" dans un sens à préciser.

Une des difficultés du problème réside dans la non-continuité des vecteurs propres de $A + \Delta$, lorsqu'ils correspondent, quand Δ tend vers 0, à une valeur propre multiple de A. Pour y remédier on utilise les sous-espaces propres ou plus généralement les sous-espaces invariants.

L'exposé se divise en trois parties ; on montre d'abord comment obtenir une décomposition spectrale approchée et partielle de $A + \Delta$ à partir d'un système complet de projecteurs sur des sous-espaces invariants de A ; on calcule ensuite des majorations pour les erreurs commises ; enfin on décrit les formules de calcul de cette décomposition spectrale approchée de $A + \Delta$, avant de les mettre en oeuvre dans un algorithme général.

1. DEFINITION ET APPROXIMATION D'UNE ISOMETRIE ENTRE LES SOUS-ESPACES INVARIANTS DE A ET CEUX DE $A + \Delta$.

1.1. Définition de l'isométrie à partir des projections orthogonales sur les sous-ensembles considérés.

On reprend d'abord un résultat que l'on restreint aux espaces de dimension finie (pour le cas général voir [4] p. 266).

Lemme (1) :

Soient E et F deux sous-espaces de l'espace vectoriel euclidien $\mathcal{E}$ de dimension finie ; on note respectivement p et q les projections orthogonales sur E et F. La condition :

$$\|p - q\| < 1 \quad (*)$$

assure que les sous-espaces E et F ont même dimension. De plus, dans ce cas,

Note :

(*) La norme matricielle utilisée est la norme subordonnée à la norme euclidienne de $\mathcal{E}$.

la transformation :

$$U = q\, H^{-1/2} p \quad \text{où} \quad H = id + p(q-p)p$$

définit une isométrie de E sur F.

Démonstration :

On note n, n_1 et n_2 les dimensions respectives de $\mathcal{E}$, E et F. On suppose que n_1 et n_2 sont différents et par exemple que $n_1 > n_2$. On en déduit que :

$$\dim(F^{\perp}) + \dim E = n - n_2 + n_1 > \dim \mathcal{E},$$

$F^{\perp}$ étant le complément orthogonal de F dans E. On est ainsi assuré qu'il existe un vecteur non nul dans l'intersection $F^{\perp} \cap E$; or ce vecteur est invariant par p-q ce qui prouve que $\|p-q\| \geq 1$.

On suppose maintenant que $\|p-q\| < 1$ et donc que $\|p(q-p)p\| < 1$. La transformation H définie dans l'énoncé est donc hermitienne, définie positive. Les transformations p et H commutent et p et $H^{-1/2}$ aussi. En remarquant que : $pH = p(id+p(q-p)p) = pqp$, on calcule la composée U^*U :

$$\begin{aligned} U^*U &= p\, H^{-1/2} q\, q\, H^{-1/2} p \\ &= H^{-1/2} p\, q\, p\, H^{-1/2} \\ &= p\, H^{-1/2}\, H\, H^{-1/2} \\ &= p \end{aligned}$$

Cela prouve que $U^*U|_E = id_E$ ce qui assure que $U|_E$ est une isométrie de E sur F. ∎

On applique maintenant cette propriété à deux systèmes complets de projecteurs orthogonaux de l'espace $\mathcal{E}$. (Par système complet de projecteurs orthogonaux, on entend une famille $(p_\ell)_{\ell=1,\ldots k}$ de projecteurs orthogonaux vérifiant :

$p_\ell p_{\ell'} = 0_{\mathcal{E}}$ lorsque $\ell \neq \ell'$ avec $\ell, \ell' = 1,\ldots,k$

et

$$\sum_{\ell=1}^{k} p_\ell = id_{\mathcal{E}}.$$

La proposition suivante est alors une conséquence directe du lemme (1) :

Proposition (2) :

On suppose que $(p_\ell)_{\ell=1,\ldots,k}$ et $(q_\ell)_{\ell=1,\ldots,k}$ sont deux systèmes complets de projecteurs orthogonaux tels que : $\|p_\ell - q_\ell\| < 1$, $\ell=1,\ldots k$. Pour tout ℓ, on note

U_ℓ la transformation définie dans le lemme (1) à partir des deux projecteurs p_ℓ et q_ℓ.

La transformation $U = \sum_{\ell=1}^{k} U_\ell$ est ainsi une isométrie qui transforme im p_ℓ en im q_ℓ pour tout ℓ.

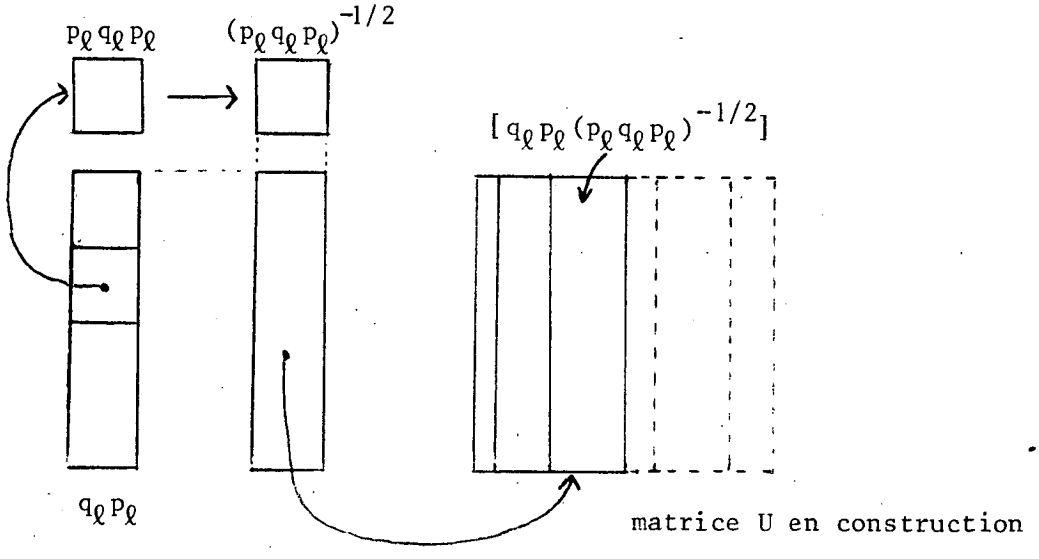
On peut remarquer que cette isométrie U peut encore s'écrire :

$$U = \sum_{\ell=1}^{k} (q_\ell p_\ell)(p_\ell q_\ell p_\ell)^{-1/2} \qquad (1)$$

en considérant $p_\ell q_\ell p_\ell$ comme un opérateur hermitien défini positif sur im p_ℓ. Donc $\mathrm{im}(p_\ell q_\ell p_\ell)^{-1/2}$ est aussi im p_ℓ. Cela entraîne que tous les composés : $(q_{\ell'} p_{\ell'})(p_\ell q_\ell p_\ell)^{-1/2}$ sont nuls lorsque ℓ et ℓ' sont différents. Ainsi on a bien la relation :

$$\sum_{\ell=1}^{k} (q_\ell p_\ell) \sum_{\ell=1}^{k} (p_\ell q_\ell p_\ell)^{-1/2} = \sum_{\ell=1}^{k} (q_\ell p_\ell)(p_\ell q_\ell p_\ell)^{-1/2}$$

Cette formulation est avantageuse pour les calculs car si on suppose que l'on exprime U dans une base orthonormée constituée de bases orthonormées des sous-espaces im p_ℓ, cette matrice U sera obtenue par juxtaposition de matrices à n lignes et dim(im p_ℓ) colonnes :



1.2. Approximation des projections.

Dans cette partie, on exprime la projection orthogonale sur un sous-espace invariant d'une matrice hermitienne M par la formule :

$$P_{M,J} = \frac{1}{2\pi i} \int_C (zI - M)^{-1} \, dz$$

où J est un intervalle de $\mathbb{R}$ ne contenant pas de valeurs propres de M sur sa frontière, et où C est le cercle de diamètre J dans $\mathbb{C}$; $P_{M,J}$ est alors la projection sur le sous espace engendré par les directions propres correspondant aux valeurs propres de M qui appartiennent à J.

On considère maintenant une matrice hermitienne A et une matrice de perturbation hermitienne Δ. On cherche alors à exprimer une relation entre $P_{A+\Delta,J}$ et $P_{A,J}$ où J est un intervalle donné.

Proposition (3) :

Soit $R(z) = (zI - A)^{-1}$ la résolvante de A ;

On considère un intervalle J ne contenant pas de valeur propre de A sur sa frontière, et C le cercle de diamètre J dans $\mathbb{C}$. On note r le rayon de ce cercle et c le maximum de $\| R(z) \|$ pour z appartenant à C. Alors, si la condition suivante est vérifiée :

$$(C1) \qquad c \, \| \Delta \| < 1$$

on peut approcher la projection $P_{A+\Delta,J}$ par la formule

$$(2) \quad P_{A+\Delta,J} = P_{A,J} + \sum_{1<n<p} \frac{1}{2\pi i} \int_C (R(z)\Delta)^n R(z) dz + R_p$$

où p est un entier donné positif

et où :

$$\| R_p \| < \frac{r \, c^{p+2} \| \Delta \|^{p+1}}{1 - c\|\Delta\|}$$

De plus, si la condition suivante est vérifiée :

$$(C2) \quad c\| \Delta\| \ (1+rc) < 1$$

alors les matrices A et $A + \Delta$ ont même nombre de valeurs propres dans J, en tenant compte de leur multiplicité.

Démonstration :

Soit z appartenant à l'ensemble résolvant de A et à C

En remarquant que l'on a alors la relation :

$$(z\, I - (A + \Delta)) = (z\, I - A)\ (I - (z\, I - A)^{-1} \Delta) \quad ;$$

on peut affirmer que la condition (C1) assure que z appartient aussi à l'ensemble résolvant de (A +Δ). La relation précédente s'écrit alors :

$$(z\ I - (A + \Delta))^{-1} = (I - R(z)\Delta)^{-1}\ R(z)$$

sous la condition (C1), la série suivante est convergente :

$$(I - R(z)\Delta)^{-1} = \sum_{n \geq 0} (R(z)\Delta)^n$$

Soit p un entier positif. De la relation

$$P_{A+\Delta,J} = \frac{1}{2\pi i} \int_C (zI - (A+\Delta))^{-1}\ dz$$

On déduit directement la formule (2) où le reste R_p est défini par :

$$R_p = \sum_{n \geq p+1} \frac{1}{2\pi i} \int_C (R(z)\Delta)^n\ R(z)\ dz$$

D'autre part :

$$\left\| \frac{1}{2\pi i} \int_C (R(z)\Delta)^n\ R(z) dz \right\| < \frac{2\pi r}{2\pi}\ c^{n+1}\ \|\Delta\|^n$$

d'où

$$\|R_p\| < \sum_{n \geq p+1} r\ c^{n+1}\ \|\Delta\|^n = \frac{r\ c^{p+2} \|\Delta\|^{p+1}}{1 - c\ \|\Delta\|}$$

Enfin pour que les matrices A et (A+Δ) aient le même nombre de valeurs propres dans J il faut et il suffit que les images des projecteurs $P_{A,J}$ et $P_{A+\Delta,J}$ aient même dimension. D'après le lemme (1), une condition suffisante pour l'assurer est que :

$$\| P_{A+\Delta,J} - P_{A,J} \| < 1$$

c'est à dire que $\| R_0 \| < 1$. Cette dernière condition s'écrit encore : $\frac{rc^2 \|\Delta\|}{1-c\|\Delta\|} < 1$; elle correspond exactement à la condition (C2). ■

2. ESTIMATION D'ERREURS

2.1. Choix des intervalles J.

On introduit les notations et hypothèses suivantes :

. les valeurs propres de A sont énumérées en ordre croissant ($\lambda_i \leq \lambda_{i+1}$, i=1,..., N-1)

. le spectre de A, $\sigma(A)$,est partitionné en k parties :

$$L_\ell = \{\lambda_{i_{\ell-1}+1}, \ldots \lambda_{i_\ell}\} \quad \text{pour } \ell = 1, \ldots, k$$

où $i_o = 0$ et $i_k = n$

. $d_{\ell,\ell+1} = \lambda_{i_\ell+1} - \lambda_{i_\ell} > 0$ est la distance entre L_ℓ et $L_{\ell+1}$

. $e_\ell = \lambda_{i_\ell} - \lambda_{i_{\ell-1}+1} \geq 0$ est l'étendue des valeurs propres dans L_ℓ

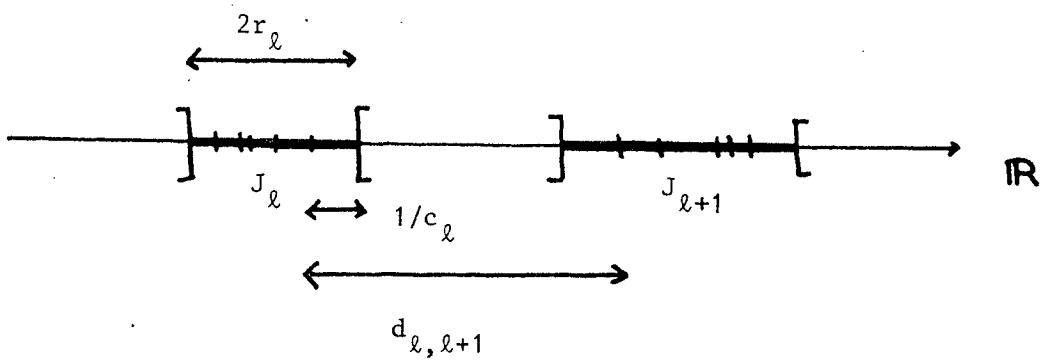
. chaque ensemble L_ℓ est inclus dans un intervalle ouvert et borné J_ℓ , ces intervalles étant disjoints deux à deux ;
le cercle de $\mathbb{C}$ de diamètre J_ℓ est noté C_ℓ et son rayon r_ℓ.

. pour tout ℓ , on note $c_\ell = \max_{z \in C_\ell} \|R(z)\| = \max_{\substack{z \in C_\ell \\ \lambda \in \sigma(A)}} \frac{1}{|z-\lambda|}$

ce maximum est atteint pour un z de J_ℓ.

. $d = \min_{\ell=1,\ldots,k-1} d_{\ell,\ell+1}$; $c = \max_{\ell=1,\ldots,k} c_\ell$; $e = \max_{\ell=1,\ldots,k} e_\ell$

$r = \max_{\ell=1,\ldots,k} r_\ell$



On suppose maintenant que la partition de $\sigma(A)$ est connue, et on cherche une distribution $(J_\ell)_{\ell=1,\ldots,k}$ d'intervalles qui minimise la quantité c.

Soit s, un entier tel que $d_{s,s+1} = d$ et a, la borne supérieure de J_s.

Suivant la position de a par rapport au milieu de cet intervalle, on peut assurer que $c_s \geqslant \frac{2}{d}$ ou $c_{s+1} \geqslant \frac{2}{d}$. Donc dans tous les cas $c \geqslant \frac{2}{d}$.

Or on peut construire une distribution d'intervalles (J_ℓ) telle que $c = \frac{2}{d}$; il suffit de prendre : $J_\ell =]\lambda_{i_{\ell-1}+1} - \frac{d}{2}, \lambda_{i_\ell} + \frac{d}{2}[$

on obtient alors $r_\ell = \frac{d + e_\ell}{2}$ et donc $r = \frac{d + e}{2}$

La condition (C2) est alors assurée pour chaque partie L_ℓ par la condition :

$$\frac{2}{d} \|\Delta\| \ (1+ \tfrac{d+e}{d}) < 1$$

que l'on peut écrire :

$$(C3) \qquad \|\Delta\| < \frac{d}{2(1+\frac{d+e}{d})}$$

ou encore :

$$(C3bis) \qquad (4 + 2\,\tfrac{e}{d})\ \|\Delta\| < d$$

Remarque : On espère que le rapport $\frac{e}{d}$ est petit, c'est à dire que les valeurs propres sont bien groupées par paquets.

2.2. Majoration de l'erreur commise sur l'isométrie

On introduit encore quelques notations ; pour tout ℓ de l'ensemble $\{1,\ldots,k\}$ on note :

. $P_\ell = P_{A,J_\ell}$; $Q_\ell = P_{A+\Delta,J_\ell}$

. $\tilde{Q}_\ell$ est l'estimation d'ordre p de Q_ℓ , définie par la relation (2)

. $V_\ell = Q_\ell P_\ell$; $W_\ell = (P_\ell Q_\ell P_\ell)^{-1/2}$; $U_\ell = V_\ell W_\ell$; $U = \sum_{\ell=1}^{k} U_\ell$

. $\tilde{V}_\ell = \tilde{Q}_\ell P_\ell$; $\tilde{W}_\ell = (P_\ell \tilde{Q}_\ell P_\ell)^{-1/2}$; $\tilde{U}_\ell = \tilde{V}_\ell \tilde{W}_\ell$; $\tilde{U} = \sum_{\ell=1}^{k} \tilde{U}_\ell$

Dans ce paragraphe, on désire calculer une majoration de $\|U - \tilde{U}\|$. Les projections

$(P_\ell)_{\ell=1,\ldots,k}$ formant un espace complet de projecteurs on obtient immédiatement que :

$$\| U - \tilde{U} \| = \max_{\ell=1,\ldots,k} \| U_\ell - \tilde{U}_\ell \|$$

On cherche donc à majorer $\| U_\ell - \tilde{U}_\ell \|$ à partir de la majoration de la quantité $\| Q_\ell - \tilde{Q}_\ell \|$. On a montré dans la proposition (3) que :

$$\| Q_\ell - \tilde{Q}_\ell \| \leqslant \delta_\ell \quad \text{avec } \delta_\ell = \frac{r_\ell c_\ell^{p+2} \| \Delta \|^{p+1}}{1 - c_\ell \| \Delta \|}$$

On démontre maintenant la proposition suivante :

Proposition (4) : Pour tout entier ℓ de l'ensemble $\{1,\ldots,k\}$ on a la majoration :

$$\| U_\ell - \tilde{U}_\ell \| \leqslant \delta_\ell \left[1 + \frac{1 + \delta_\ell}{(1-\rho_\ell)^2} \right]$$

où $\rho_\ell = \dfrac{r_\ell c_\ell^2 \| \Delta \|}{1 - c_\ell \| \Delta \|} < 1$ (condition (C2)).

Démonstration : Pour obtenir la majoration cherchée, on utilise l'inégalité suivante :

$$\| U_\ell - \tilde{U}_\ell \| \leqslant \| V_\ell - \tilde{V}_\ell \| . \| W_\ell \| + \| V_\ell \| . \| W_\ell - \tilde{W}_\ell \|$$

Pour alléger l'écriture, on omet l'indice ℓ dans la suite de la démonstration. Sur les quatre normes à majorer, trois s'obtiennent immédiatement :

$$\| V - \tilde{V} \| = \| (Q - \tilde{Q}) P \| \leqslant \| Q - \tilde{Q} \| \leqslant \delta$$

$$\| \tilde{V} \| \leqslant \| V \| + \| V - \tilde{V} \| \leqslant 1 + \delta$$

$$\| W \| = \| (P Q P)^{-1/2} \| = (\| P Q P \|)^{-1/2} \leqslant 1$$

Il reste maintenant à majorer la quantité $\| W - \tilde{W} \|$. Pour cela on utilise le lemme suivant :

lemme 5 : soit $f(x) = \sum_{n>o} a_n x^n$ une série entière de rayon de convergence égal à 1, telle que $\sup_{n \in \mathbb{N}} |a_n| = a < + \infty$

Soient A et B deux matrices hermitiennes et telles que :

$$\rho = \max (\rho(A), \rho(B)) < 1.$$

Alors : $\| f(A) - f(B) \| < \dfrac{a}{(1-\rho)^2} \| A - B \|$

Démonstration du lemme : On part de la relation suivante :

$$f(A) - f(B) = \sum_{n\geqslant 1} a_n (A^n - B^n)$$

Soit d_n la norme de $A^n - B^n$. Pour majorer cette quantité on utilise l'identité suivante :

$$A^n - B^n = (A - B)A^{n-1} + B(A^{n-1} - B^{n-1}) ,$$

afin d'écrire l'inégalité :

$$d_n \leqslant d_1 \rho^{n-1} + \rho d_{n-1}$$

Par récurrence, on obtient alors la majoration :

$$d_n \leqslant d_1 n \rho^{n-1}$$

La majoration obtenue pour la norme de $f(A) - f(B)$ est alors :

$$\| f(A) - f(B) \| \leqslant a d_1 \sum_{n\geqslant 1} n \rho^{n-1} = \frac{a d_1}{(1-\rho)^2} \quad \blacksquare$$

fin de la démonstration de la proposition :

On applique le lemme à la fonction $f(x) = (1+x)^{-1/2}$ qui se développe en série entière au voisinage de 0 :

$$f(x) = \sum_{n\geqslant 0} a_n x^n \quad \text{avec} \begin{cases} a_n = (-1)^n \dfrac{1.\,3.\,\ldots.(2n-1)}{2.\,4.\,\ldots.\,2n} , \; n \geqslant 1 \\ \text{et} \\ a_o = 1 \end{cases}$$

le rayon de convergence de cette série est 1, et $a = \sup_{n\geqslant o} |a_n| = 1$

Les matrices A et B du lemme seront ici les matrices $(P\,Q\,P - P)$ et $(P\,\tilde{Q}\,P - P)$. Elles sont hermitiennes, de rayon spectral inférieur à 1 ; en effet on peut facilement montrer que

$\| P\,Q\,P - P \|$ et $\| P\,\tilde{Q}\,P - P \|$ sont toutes deux majorées par le majorant

trouvé pour $\|R_o\|$ c'est à dire $\frac{r\,c^2\,\|\Delta\|}{1 - c\|\Delta\|}$ qui est inférieur à 1 lorsque la condition (C2) est vérifiée.

D'autre part, on a les relations suivantes :

$$f(P\ Q\ P - P) = (P\ Q\ P)^{-1/2}$$

$$f(P\ \tilde{Q}\ P - P) = (P\ \tilde{Q}\ P)^{-1/2}\ ,\ \text{car P représente l'identité sur im P}$$

On en déduit

$$\| W - \tilde{W} \| \leqslant \frac{\delta}{(1-\rho)^2}$$

En conclusion, on a obtenu la majoration,

$$\| U_\ell - \tilde{U}_\ell \| \leqslant \delta_\ell .\ 1 + (1 + \delta_\ell)\ \frac{\delta_\ell}{(1-\rho)^2}$$ ■

Corollaire (6) :

On a les estimations suivantes :

$$\| U - \tilde{U} \| = O\ (\Delta^{p+1})$$

$$\tilde{U}^*\ \tilde{U} = I + O\ (\Delta^{p+1})$$

La première estimation est une conséquence directe de la proposition (4). On peut donc écrire $U = \tilde{U} + E$ avec $E = O\ (\Delta^{p+1})$

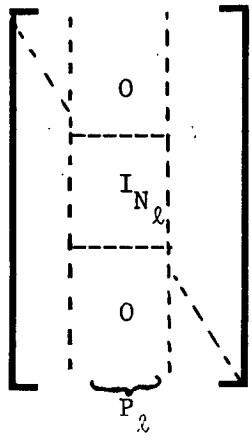
$$\tilde{U}^*\ \tilde{U} = (U - E)^*\ (U - E) = I - E^*\ U - U^*\ E + E^* E$$ ■

III - CALCUL EFFECTIF DE L'ISOMETRIE

3.1. Calcul des projections

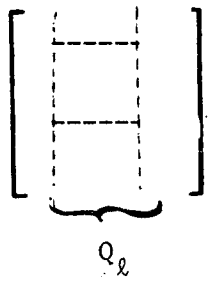
Les calculs de ce paragraphe sont réalisés en supposant que la matrice A est diagonale, ce qui correspond à avoir effectué le changement de base de matrice X. On calcule ici l'estimation d'ordre 2 de la projection $Q = P_{A+\Delta, J_\ell}$. En fait la relation (1) assure qu'il suffit de calculer la matrice QP où $P = P_{A, J_\ell}$, puisqu'ensuite on calculera la matrice $QP\ (P\ Q\ P)^{-1/2}$ et cela pour chaque intervalle J_ℓ.

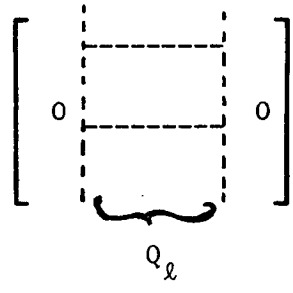
On remarque d'abord que P peut s'écrire :

$P =$ 

où N_ℓ = dim. im P

La matrice QP aura la forme suivante :



$P =$ 

Il suffit donc de calculer la matrice rectangulaire $\tilde{Q}_\ell$ de type N x N_ℓ , qui approche à l'ordre 2 la matrice Q_ℓ. Cette matrice peut s'écrire

$$\tilde{Q}_\ell = P_\ell + T_\ell^{(1)} + T_\ell^{(2)}$$

En reprenant les notations du paragraphe 2-1, on sait que les valeurs propres de A appartenant à J_ℓ sont $\lambda_{i_{\ell-1}+1}$, ... , λ_{i_ℓ}

(donc $N_\ell = i_\ell - i_{\ell-1}$) .

La matrice Δ est de terme général δ_{ij}.

Soit z un élément de l'ensemble résolvant de A, alors

$$(\Delta R(z))_{ij} = \frac{\delta_{ij}}{z - \lambda_j}$$

On partage l'ensemble I = {1,...,N} en trois parties : $I_1 = \{1,\ldots,i_{\ell-1}\}$

$I_2 = \{i_{\ell-1}+1,\ldots,i_\ell\}$

$I_3 = \{i_\ell + 1,..,N\}$

3.1.1 : terme du 1er ordre

$$T_\ell^{(1)} = \left(\frac{1}{2\pi i} \int_{C_\ell} R(z)\,\Delta R(z)\,dz\right) P_\ell$$

$$= \frac{1}{2\pi i} \int_{C_\ell} \left(R(z)\,\Delta R(z)\,P_\ell\right) dz$$

Or $$\left[R(z)\,\Delta\,R(z)\right]_{ij} = \frac{\delta_{ij}}{(z-\lambda_i)(z-\lambda_j)}$$

Il suffit de calculer ce terme pour $j \in I_2$ dans ce cas on trouve que :

$$\left[T_\ell^{(1)}\right]_{ij} = \begin{cases} \dfrac{\delta_{ij}}{\lambda_i - \lambda_j} & \text{pour } i \in I_1 \cup I_3 \\ 0 & \text{pour } i \in I_2 \end{cases}$$

Remarque : Si on note $T^{(1)}$ l'assemblage des matrices $T_\ell^{(1)}$:

$$T^{(1)} = \left[T_1^{(1)}, \ldots, T_k^{(1)}\right]$$

on remarque que $T^{(1)}$ est antihermitienne. Il suffira donc de calculer $\left[T_\ell^{(1)}\right]_{ij}$ pour $i \in I_1$.

3.1.2 : terme du 2e ordre

$$T_\ell^{(2)} = \frac{1}{2\pi i} \int_{C_\ell} \left[R(z)\,\Delta R(z)\,\Delta R(z)\,P_\ell\right] dz$$

Or $$\left[R(z)\,\Delta R(z)\,\Delta R(z)\right]_{ij} = \sum_{u=1}^{N} \frac{\delta_{iu}\,\delta_{uj}}{(z-\lambda_i)(z-\lambda_u)(z-\lambda_j)}$$

On montre facilement que :

$$\frac{1}{2\pi i} \int_C \frac{1}{(z-a)(z-b)(z-c)}\,dz = \begin{cases} 0 \text{ si a, b, c, extérieurs à D} \\ \dfrac{1}{(c-a)(c-b)} \text{ si a et b extérieurs et c intérieur à D} \\ -\dfrac{1}{(c-a)(c-b)} \text{ si a et b intérieurs et c extérieur à D} \\ 0 \text{ si a, b, c, intérieurs à D} \end{cases}$$

C : contour d'un domaine D fermé

d'où :

$$\text{pour } j \in I_2 : (T_\ell^{(2)})_{ij} = \begin{cases} \sum_{u \in I_1 \cup I_3} \frac{\delta_{iu}\,\delta_{uj}}{(\lambda_j-\lambda_i)(\lambda_j-\lambda_u)} - \sum_{u \in I_2} \frac{\delta_{iu}\,\delta_{uj}}{(\lambda_j-\lambda_i)(\lambda_u-\lambda_i)} & \\ \qquad\qquad \text{si } i \in I_1 \cup I_3 & \\ \sum_{u \in I_1 \cup I_3} \frac{\delta_{iu}\,\delta_{uj}}{(\lambda_u-\lambda_i)(\lambda_j-\lambda_u)} \quad \text{si} \quad i \in I_2 & \end{cases}$$

3.2. Calcul de $H^{-1/2}$ lorsque H est une matrice hermitienne définie positive

Dans le calcul de l'isométrie, il est nécessaire de calculer la puissance $-\frac{1}{2}$ de blocs diagonaux. Même si on espère que ceux ci sont de petite taille par rapport à N, il n'est pas souhaitable d'avoir à les diagonaliser, car cela pénaliserait l'algorithme. On recherche donc une formule de récurrence de la forme suivante :

$$n \geqslant 0 \quad T_{n+1} = P(T_n, H)$$

où P est un polynôme à deux variables. Si T_n et H sont hermitiennes et admettent une base commune de vecteurs propres alors il en est de même pour les matrices T_{n+1} et H, car pour cela il faut et il suffit que T_n et H soient hermitiennes et commutent. Afin d'assurer cette propriété au premier rang on peut choisir $T_o = I$.

On suppose que Y est une matrice unitaire qui diagonalise simultanément T_n et H :

$$H = Y \Lambda_n Y^* \quad \text{et } T_n = Y\ D_n\ Y^*, \quad \text{où } \Lambda_n \text{ et } D_n \text{ sont diagonales.}$$

La relation $D_{n+1} = P(D_n, \ )$ assure ainsi que l'on peut raisonner sur chaque composante. Il suffit donc de trouver un polynôme P tel que le schéma :

$$\begin{cases} d_{n+1} = P(d_n, \lambda) \quad n \geqslant 0 \quad \text{avec} \quad \lambda > 0 \\ do = 1 \end{cases}$$

définisse une suite (d_n) qui converge vers $\frac{1}{\sqrt{\lambda}}$.

Le polynôme $P(d,\lambda) = \frac{d}{2}(3 - d^2 \lambda)$ répond au problème lorsque $\lambda \in]0,3[$. En fait on ne l'utilisera que sur l'intervalle $]0,2[$.

Proposition (7) :

Soit $H = I + R$ où R est une matrice hermitienne telle que $\|R\| < 1$.

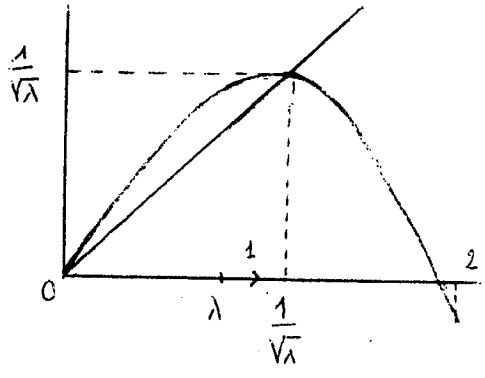
Alors le schéma : $\begin{cases} T_o = I \\ T_{n+1} = \frac{1}{2} T_n (3I - T_n^2 H) \end{cases}$

définit une suite de matrices hermitiennes (T_n) commutant avec H et qui admettent pour limite la matrice $H^{-1/2}$.

Ce schéma est d'ordre 2.

Soit $f(x) = \frac{x}{2}(3 - x^2 \lambda)$.

Sur l'intervalle $[0,2]$ la fonction a le sens de variation suivant :



Pour $\lambda \in]0,2[$:

$f(1) = \frac{1}{2}(3 - \lambda) > 0$

donc $d_1 \in]0, \frac{1}{\sqrt{\lambda}}[$.

comme $f(]0, \frac{1}{\lambda}[) =]0, \frac{1}{\lambda}[$

on en déduit que pour tout $n \geq 1$ on a $d_n \in]0, \frac{1}{\sqrt{\lambda}}[$

Soit $\Delta_n = \frac{1}{\sqrt{\lambda}} - d_n$. Alors :

$$\Delta_{n+1} = \frac{3}{2}\left(\frac{1}{\sqrt{\lambda}} - d_n\right) - \frac{\lambda}{2}\left(\left(\frac{1}{\sqrt{\lambda}}\right)^3 - d_n^3\right)$$

$$= \left[1 - \frac{\lambda}{2}\left(d_n^2 + \frac{1}{\sqrt{\lambda}} d_n\right)\right] \Delta_n$$

or lorsque $d_n \in]0, \frac{1}{\sqrt{\lambda}}[$ on a $0 \leq 1 - \frac{\lambda}{2}(d_n^2 + \frac{1}{\sqrt{\lambda}} d_n) < 1$,

ce qui assure la convergence du schéma.

Ce schéma est d'ordre 2 car :

$$\Delta_{n+1} = \frac{\lambda}{2} \left(d_n + \frac{2}{\sqrt{\lambda}}\right) \Delta_n^2$$

d'où $$|\Delta_{n+1}| \leqslant K \Delta_n^2 \quad \text{avec } K = \frac{3}{2} \sqrt{\lambda}$$

Pour la matrice initiale T_o, on peut prendre à la place de I l'estimation d'ordre k (k étant un nombre fixé) du développement de $(I + R)^{-1/2}$ dans la série de Taylor.

Soit $T_n = T + \Delta_n$; la proposition précédente montre que $\Delta_n = O(\Delta_0^{2^n})$ et donc qu'avec la matrice initiale T_o choisie on aura $\Delta_n = O(R^{(k+1)2^n})$

Cette précision sera obtenue au prix de (k-1) produits de matrices pour calculer T_o et de 3n produits de matrices pour calculer T_n.

Soit $\phi(k,n) = (k+1)2^n$, l'exposant de R dans l'ordre de l'estimation de Δ_n, et $M(k,n) = (k-1) + 3n$ le nombre de produits de matrices nécessaires pour y parvenir. On cherche alors à minimiser la fonction M, pour une valeur de f donnée. En se plaçant dans $\mathbb{R}_+^2$ le problème devient :

$$\begin{cases} \text{sous la contrainte} & \phi(x,y) = a \\ \text{minimiser } M(x,y) \end{cases}$$

On trouve que le minimum est atteint pour $x_o = \frac{3}{\text{Log } 2} - 1 \simeq 3{,}3$ et $y_o = \log_2\left(\frac{a}{x_o+1}\right)$

La valeur k sera donc 3 ou 4.

Un petit raisonnement permet de montrer que si on choisit k = 3, ce choix est optimal avec une probabilité de $\frac{3}{5}$. S'il n'est pas optimal on aura effectué deux produits de matrices en trop.

Dans l'algorithme on choisit donc k = 3, c'est à dire :

$$T_o = I - \frac{1}{2} R + \frac{3}{8} R^2 - \frac{15}{48} R^3$$

que l'on calcule sous la forme :

$$T_o = \left(\left(-\frac{15}{48} R + \frac{3}{8} I\right) R - \frac{1}{2} I\right) R + I$$

On remarquera aussi que tous les produits de matrices effectués le sont sur des matrices hermitiennes et commutant. Le produit est ainsi hermitien lui aussi, ce qui réduit les calculs de moitié environ.

3.3. Algorithme général

On suppose que l'on désire diagonaliser une matrice A' hermitienne et qu'on connait une matrice unitaire X qui approche une matrice de vecteurs propres de A'. Soit D la matrice diagonale obtenue par :

$$D = \text{diag}\,(X^* A'X) \text{ et soit } \Delta' = X^* A'X - D$$

On en déduit donc que :

$$A' = A + \Delta \qquad \text{où } A = X\ D\ X^*$$
$$\text{et } \Delta = X\ \Delta'\ X^*$$

on est ainsi dans le cadre de l'étude. Soit U une matrice unitaire de vecteurs propres de $D + \Delta'$ alors XU est une matrice unitaire de vecteurs propres de A'.
En fait cet algorithme est employé itérativement, jusqu'à obtenir la précision cherchée dans la décomposition spectrale de A'.

Sa structure est la suivante :

données : A' : matrice à diagonaliser
X : matrice unitaire qui approche une matrice de vecteurs propres de A'
ε : paramètre de précision.

Boucle :

calcul de la matrice réduite : $\Delta = X^* A'X$
calcul de la norme de la partie hors diagonale de Δ

tant que (norme $\geq \varepsilon$)
seuil := 4 * norme

(1) groupement

Pour chaque tranche :
faire
(2) calcul de la norme de la partie extérieure au bloc diagonal
si(largeur tranche>1)alors diagonalisation partielle du bloc diag. fsi
choix de l'ordre de la correction de la tranche.
fait

(3) Correction des tranches

Orthonormalisation de X.

Fin Boucle

On détaille maintenant les trois parties encadrées de l'algorithme.

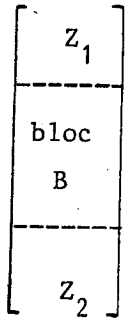
Groupement (1)

On note $(\lambda_i)_{1,N}$ la diagonale de Δ ; cette partie de l'algorithme vérifie que cette suite de valeurs est croissante au seuil de discernement près :
pour $j \geqslant 1$: $\lambda_{i+j} \geqslant \lambda_i$ - seuil

Si ce n'est pas le cas, on réordonne la suite en effectuant les permutations correspondantes sur Δ et X.
Ensuite, on regroupe les valeurs (λ_i) en paquets dans lesquels la distance entre deux valeurs consécutives est inférieure au seuil, tandis que ces paquets sont distants les uns des autres d'une distance supérieure à ce seuil.

Diagonalisation des blocs (2)

* Pour un groupe on considère la tranche correspondante dans X^* A' X :



où $\rho = \max (\| Z_1 \|', \| Z_2 \|')$

$\| M \|' = \sup_{i,j} |m_{ij}|$

On applique à B une étape de l'algorithme de jacobi à tous les éléments de B dont le module est supérieur à ρ.
On choisit maintenant l'ordre de correction de la tranche :

si $\rho < \varepsilon$ alors l'ordre de correction est 0 (pas de correction).
si $\varepsilon < \rho < \varepsilon^{2/3}$ alors l'ordre est 1.
si $\varepsilon^{2/3} < \rho$ alors l'ordre est 2.

correction des tranches :

Cette partie correspond à l'étude de l'article. On y calcule une matrice U de vecteurs propres approchés de Δ. On effectue alors le produit $X := X U$.

CONCLUSION

A partir de l'algorithme présenté ici, on a écrit un code (dans le cas d'une matrice symétrique réelle) qui a permis de vérifier les résultats établis dans l'exposé. Il reste maintenant à reprendre ce programme pour y minimiser les calculs et en faire un code compétitif avec la meilleure méthode (tridiagonalisation et QL). En effet dans le cas où il est nécessaire de diagonaliser complètement une matrice hermitienne qui dépend d'un paramètre, en une suite de valeurs de ce paramètre, on espère aboutir à une méthode plus rapide que celles qui ne tiennent pas compte de la diagonalisation de l'étape précédente.

BIBLIOGRAPHIE

[1] Chandler Davis and W.M. Kahan - The rotation of Eingenvectors by a perturbation III - SIAM Journal on Numer. Anal., (Mar.70), vol.7, n°1, pp. 1-46.

[2] F. Chatelin-Laborde - Perturbation d'une matrice hermitienne ou normale - Numer. Math., (1971), 17, pp. 318-337.

[3] T. Kato - Perturbation theory for linear operators - Springer-Verlag (1966).

[4] F. Riesz et B.Sz. Nagy - Leçons d'analyse fonctionnelle - Gauthier-Villars (1968).

[5] G.W. Stewart - Error and perturbation bounds for subspaces associated with certain Eingenvalue problems - SIAM Review, (Oct.73), vol.15, n°4, pp. 727-764.

[6] G.W. Stewart - On the perturbation of pseudo-inverses, projections and linear least squares problems - SIAM Review, (Oct.77), vol.19, n°4, pp. 634-662.

[7] J.H. Wilkinson - The algebraic eingenvalue problem - Clarendon Press (1965).

PI 208 Problèmes d'implémentation du langage Prolog en vue de la réalisation d'une machine Prolog
Yves BEKKERS, Bernard CANET, Olivier RIDOUX, Lucien UNGARO
Octobre 1983, 63 pages.

PI 209 La technique du suivi de contour en synthèse d'images et ses applications
Gérard HEGRON, Octobre 1983.

PI 210 A new characterization of infinitary rational languages
Philippe DARONDEAU, Laurent KOTT
Octobre 1983, 9 pages.

PI 211 On the observational semantics of fair parallelism
Philippe DARONDEAU, Laurent KOTT
Octobre 1983, 40 pages.

PI 212 Solution à forme produit d'un système linéaire
J. PELLAUMAIL
Novembre 1983, 36 pages.

PI 213 Equations de Chapman-Kolmogorov et flots stationnaires pour des processus markoviens
J.Y. LE BOUDEC et J. PELLAUMAIL
Novembre 1983, 18 pages.

PI 214 Distribution des interentrées et intersorties pour des réseaux à forme produit
J.Y. LE BOUDEC
Novembre 1983, 39 pages.

PI 215 Un outil informatique pour l'analyse graphique de données
René THORAVAL
Juin 1983

PI 216 Perturbation de la décomposition spectrale d'une matrice hermitienne
Bernard PHILIPPE
Novembre 1983, 21 pages.

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique